Assessment of annoyance, loudness and unpleasantness with different recording and playback techniques

Celik, Emine; Persson Waye, Kerstin; Møller, Henrik

Published in:
The 2005 Congress and Exposition on Noise Control Engineering

Publication date:
2005

ENVIRONMENTAL NOISE CONTROL
The 2005 Congress and Exposition on Noise Control Engineering
07-10 August 2005 - Rio de Janeiro - Brazil

Assessment of annoyance, loudness and unpleasantness with different recording/playback techniques

Emine Çelik[a], Kerstin Persson Waye[b], Henrik Møller[c]

[a,c]Department of Acoustics, Aalborg University, Fredrik Bajers Vej 7 B5, DK-9220, Aalborg Ø, Denmark
[b]Department of Environmental Medicine, The Sahlgrenska Academy, Göteborg, Sweden

[a]emc@acoustics.aau.dk; [b]kerstin.persson-waye@envmed.gu.se; [c]hm@acoustics.aau.dk

Abstract The present study was carried out to investigate if there is a difference in perception of annoyance, loudness and unpleasantness between monophonic recordings played back through a loudspeaker and binaural recordings played back via headphones, and to evaluate whether a possible difference depends on temporal and frequency characteristics as well as spatial characteristics of the sounds. Evaluations were also done in order to see a possible effect of durations of sound exposures. The experiment adopted three psychometric methods for achieving responses from subjects. Fifty-four young students participated and three types of sounds were used in the experiments: everyday "restaurant" sound (from using cutlery at platters, moving chairs, talking etc.), road traffic sound and a low-frequency ventilation sound. The sounds were recorded with two different techniques (monophonic and binaural). The monophonic recordings were presented through a loudspeaker and the binaural recordings were presented through both closed (circum-aural) and completely open (free-of-the-ear) headphones. Each sound was played back at three different levels. The results show that for all judgments (annoyance, loudness and unpleasantness), there was no significant main effect of recording and playback techniques; however significant interactions between techniques and sounds were found. For annoyance and unpleasantness, an influence of psychoacoustic method was found.

1. INTRODUCTION

Our auditory and perceptual response to a sound is multidimensional and more complex than what can be evaluated by a simple frequency weighted, time averaged measure such as equivalent A-weighted sound pressure levels [1], [6] and [5]. Although, earlier research on psychoacoustics [8] has resulted in refined measures, with the exception of loudness, few studies have been able to convincingly link these measures to annoyance. On the contrary, these measures were not sensitive enough to differentiate between annoyances, even for sounds of similar origin [4]. In order to study the importance of sound properties on annoyance, further experimental studies need to be undertaken. Experimental studies on noise annoyance are often criticized on the grounds that conclusions based on experimental data cannot be applied to settings outside the laboratory, because the experimental conditions are too unlike the real world. In order to ensure validity between real life and experimental settings several conditions related to recording, playback and context of the experimental situation needs to be attended to. In previous studies of perception and response to sounds, several methods have been adopted both with regard to recording techniques (monophonic or binaural), playback techniques (through headphones or loudspeakers) and subjective evaluation techniques. Regarding recording and playback techniques very little is known on how these techniques affect the subjective perception and overall response. A better knowledge in this field is crucial in order to compare sound exposures between studies.

A major difference between the two recording and playback techniques is their ability to reproduce spatial properties of the sound. A further difference exists for low frequencies, which at higher sound pressure levels do not only affect the hearing but also give sensations in other parts of the body, mainly the chest and abdomen [3].

The purpose of this study is to investigate whether there is a difference in subjective perception and response related to annoyance, loudness and unpleasantness between mono recordings, played back through a loudspeaker, and binaural recordings played back via headphones. A further aim was to evaluate whether the perception differed depending on temporal, spectral and/or spatial characteristics of the sound. The study also adopts various psychometric methods for achieving responses from subjects. Many of the response methods used today are based on short-term comparisons of sounds and it can be questioned whether they can be used to measure annoyance or even unpleasantness. Therefore the project also aimed to evaluate the effect of duration on assessment of annoyance, loudness and unpleasantness.

2. MATERIAL AND METHODS

2.1 Subjects

A total of 54 paid native Danish speaking volunteers participated in the experiments (27 females and 27 males aged between 20 and 34 years, M=24.72, SD=2.78). The subjects had not previously participated in similar sound evaluation experiments. Audiometric tests [ISO 8253-1] ensured normal hearing within 15 dB at the octave band frequencies 125 Hz to 4 kHz and 20 dB at 8 kHz (Madsen Orbiter 922 audiometer, automatic mode-ascending method). To assess the subjects' noise sensitivity in general, a questionnaire [7] translated into Danish was answered after the audiometric tests. The questionnaire had a total of 120 points; the higher the point scores, the higher sensitivity to noise. The subjects' answers ranged between 48 and 111 points with an average of 72.5 (SD=11.83). The subjects were allocated to three groups that judged different basic psychoacoustic attributes; annoyance, loudness and unpleasantness. Females and males were separately ordered on the basis of their noise sensitivity scores and were then randomly distributed into the three groups.

2.2 Sounds

Three sounds were used in the study. The sounds varied in particular with regard to spatial properties and content of low frequencies (20-200 Hz). The recordings were done with a Harmonie 01 dB system using an artificial head [2] for the binaural recordings and a G.R.A.S 40 EN microphone for the monophonic recordings. The first sound (R) comprised sounds typically occurring in a restaurant. Sounds from using cutlery at platters, moving chairs and people talking occurred in all directions in the original sound field. The conversations were done in Turkish (female voice) and Spanish (male voice in Costilla La Mancha accent) so the conversation would be meaningless to the test subjects. The second sound, traffic sound (T), was obtained from a road in front of the listener and thus sound sources occurred in a limited spatial range in the original sound field. The third sound, ventilation sound (V), was recorded in a large basement room with ventilation channels, and there was no obvious direction to the sound source(s). In order to obtain a predominantly low frequency character, sound pressure levels in the frequency region of 31.5 to 125 Hz were increased during data processing. Each sound was recorded for approximately 2 minutes (binaurally and monaurally) and these recordings were used to prepare the experiment sounds, which were 5 seconds and 10 minutes. Care was taken to prepare the 5-seconds sounds so that they were representative of the 10-minute sounds. Each sound was reproduced at 3 different levels: naturally occurring level at the recording time (0 dB), 6 dB below (-6 dB) and 6 dB above (+6 dB). For ventilation sound the low-frequency boosted version is referred to as the natural level. The equivalent A-weighted sound pressure levels (L_{Aeq}) of 10 minute sounds ranged from 52 to 59 dB while 5 second sounds ranged from 51 to 55 dB (natural level).

2.3 Exposure room and playback setup

The experiments were carried out in a room (l=8.10 m, w=6.96 m, h=3.05 m, V=172 m^3), which was partly furnished as a living room with a two-person sofa, two armchairs, a small table, and some plants. In order to try to overcome some of the hinders for generalizing the data to the real environment, extra effort was spent to keep the experimental setup as close as possible to the real life situation, in a controlled environment. Figure 1 shows the listening test set up. The sound pressure level (SPL) of the background noise (including the room ventilation and cooling system) was below the hearing threshold levels [ISO 226; 2003] for every 1/3 octave frequency band in the 25 Hz-12.5 kHz range.

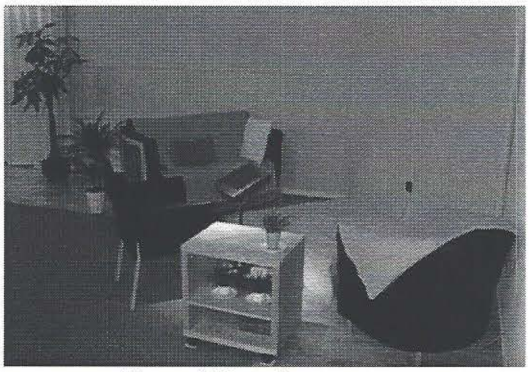

Figure 1 Listening test setup

The monophonic recordings were presented through a loudspeaker system (Genelec 1031A/1094A) (technique ML), which was hidden behind a curtain, and the binaural recordings were presented through either circum-aural headphones (Beyerdynamics DT 990) (technique BH1) or headphones that were completely open and free of the ear (AKG K 1000) (technique BH2). In technique BH2, taking the limitations (harmonic distortion during the playback of low frequency sound) of the open headphone into account, it became inevitable to play back the sound in a different way than normal binaural playback. The low frequency part (lower than 100 Hz) was reproduced through the loudspeaker and the rest through the open headphone, so the subjects were fully exposed to the low-frequency sound field without loosing the spatial perception connected to the binaural technique.

2.4 Evaluation methods

In Method I, each group rated either annoyance, loudness or unpleasantness by answering the question: How XX did you find the sound?. (XX was replaced by annoying, loud and unpleasant for the three different groups. The answers were given on 100 mm horizontal scales with the anchor points '*not at all ....*' and '*very ......*' on an electronic tablet after each exposure, transformed to a number between 0 and 100 and automatically stored on a computer. Degree of annoyance, loudness or unpleasantness was measured in mm. In Method II (paired comparisons) the subjects made forced-choice paired comparisons of annoyance, loudness or unpleasantness (depending on the group) of sounds. The two sounds in a pair were presented with a 1-s pause between. The questions were posed: "Which of the sounds were you more annoyed by?", "Which of the sounds did you find louder?" or "Which of the sounds did you find more unpleasant?" The answers were given on an electronic tablet after each exposure, where one of two alternatives had to be chosen. The sounds in a pair were either from the same technique or from different techniques. Only ML and BH2 techniques were used in this session, since only these would allow comparisons across techniques without the need of taking the headphones on and off between the two sounds in a pair. Method III was semantic description method which results will be reported elsewhere.

2.5 Experimental design and procedure

For Method I the study had for each group a 3 (sounds) × 3 (levels) × 3 (techniques) × 2 (durations) factorial design with repeated measures. The long stimuli were given on separate days with one technique per day, and subjects were asked to choose a book out of 5 alternatives and read it during the test. The short stimuli were given on one day, and in order to allow an evaluation of the subjects' reliability all stimuli appeared twice. The order of techniques (ML, BH1 and BH2) was balanced between subjects (same order for long and short experiments). The order of stimuli was randomized for each subject, technique and duration.

In Method II, 18 stimuli were included (2 techniques × 3 sounds × 3 levels). The pairs were taken from a half matrix design that excludes identical and reverse pairs, thus giving a total of 153 pairs ($n*(n-1)/2$, $n=18$). The order of the pairs was randomized. With the given design, for comparisons within the same technique, each sound/level combination occurred once with any other sound/level combination, and for these the order of the sound/level combination was random. For across techniques comparisons, each sound/level combination occurred twice with any other sound/level combination. The first time the order of the techniques was random, while it was reversed the second time. On a separate day before the experiment (preparation day), subjects underwent an audiometric test and filled in the noise sensitivity questionnaire. Each subject took part in sessions on six separate days (with a minimum of 48 hours in between) and always at the same time of the day. During all sessions subjects were given breaks at regular intervals, in order to avoid tiredness. For each group half of the subjects completed Method I-short and Method II on their first experimental day, while the other half started with the three days of Method I-long. For all subjects, Method III was carried out on their last day. The experimental schedule for each group is given in Table 1.

Prior to each method and technique, subjects were given written and verbal instructions, and they listened to 5 seconds of each sound in all level in a random order. They also underwent a learning session in order to get familiar to the test method. The subjects were instructed to remain seated in the same position, (upright leaning against the back of the chair and without moving their head) throughout the test. They were also informed that during the test they would be supervised by the operator (by mean of intercom and camera). Subjects were instructed to give their immediate response.

Table 1 The experimental schedule for each of the three 18-subject groups (A,L,U)

Day	9 subjects	9 subjects
Preparation day	Audiometry test & Noise Sensitivity Questionnaire	Audiometry test & Noise Sensitivity Questionnaire
1. Day	Method I-short + Method II	Method I-long
2. Day	Method I-long	Method I-long
3. Day	Method I-long	Method I-long
4. Day	Method I-long	Method I-short + Method II
5. Day	Method III	Method III
TOTAL	Preparation=35 min Experiment=351 min	Preparation=35 min Experiment=351 min

2.6 Analysis and statistical methods

For Method I-short, where there were repetitions, subject reliability was checked by a two-way random effects model with intra-class correlation coefficients using an absolute agreement definition. For each group analyzes of variance (ANOVA) with repeated measures were performed to evaluate the influence of technique, sound, level and duration as well as interactions between these. The p-values are based on degrees of freedom, corrected with Greenhouse-Geisser epsilon for sphericity, when appropriate. A one-way ANOVA was performed to evaluate the difference between three psychoacoustic attributes. The statistical analyses were carried out using SPSS. All tests were two-tailed, and a p-value below 0.05 was considered statistically significant (mean difference is abbreviated as MD and 95% confidence interval as 95% CI)

3. RESULTS

3.1 Annoyance

No significant main effects of technique and duration on annoyance ratings were found.
A significant main effect of sound (($F(1.342,17)=26.508$, $p=0.000$) was found. Restaurant sound was significantly different from traffic (MD=13.8, 95% CI=(8.2-19.5)) and ventilation sounds (MD=16.1, 95% CI=(7.9-24.4)). Traffic and ventilation sounds were not significantly different (Figure 2).
A significant main effect of level (($F(1.223,17)=65.374$, $p=0.000$) was found. All levels were significantly different from each other (0 dB versus -6 dB: MD=8.6, 95% CI=(5.2-11.8); +6 dB versus 0 dB: MD=9.7, 95% CI=(6.4-13.0); +6 dB versus -6 dB: MD=18.3, 95% CI=(12.6-24.0)) (Figure 2).

A significant two-way interaction was found between technique and sound ((F(4,17)=6.921, p=0.000). For ML technique, restaurant sound was judged less annoying, traffic and ventilation sounds were judged more annoying.

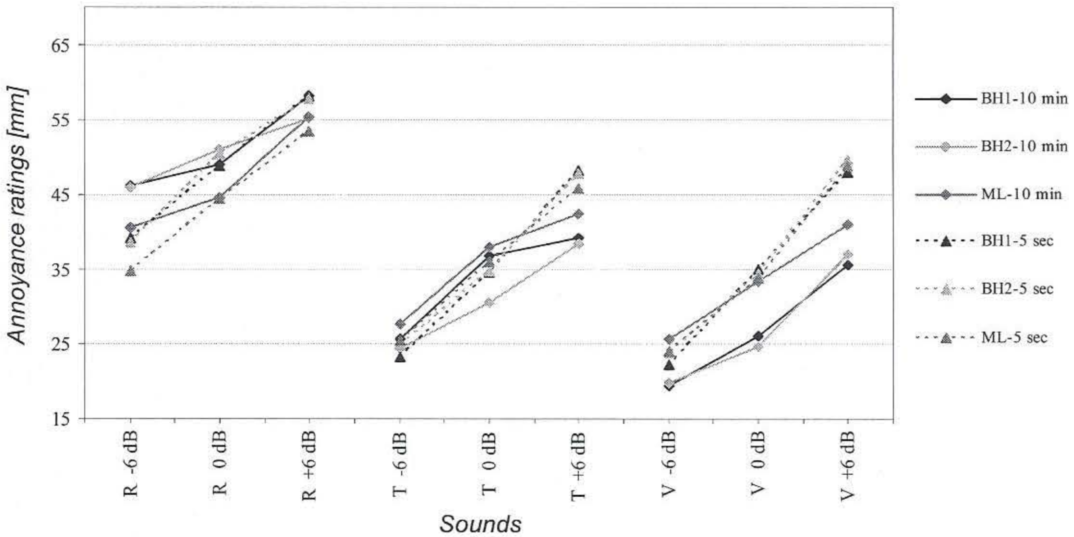

Figure 2 The mean values of annoyance ratings for 18 subjects for all combinations of sound, level, technique and duration.

A significant two-way interaction was also found between sound and duration ((F(2,17)=5.907, p=0.006). Restaurant sound was judged more annoying during 10-minute exposure, traffic and ventilation sounds were judged more annoying during 5-second exposure.

A significant two-way interaction was found between level and duration ((F(2,17)=10.222, p=0.000). -6 dB level was judged more annoying during 10-minute exposure but 0 dB and +6 dB levels were judged more annoying during 5-second exposure. The mean difference between annoyance judgements for two exposure durations was highest at +6 dB level.

3.2 Loudness

No significant main effects of technique and duration on loudness ratings were found.

A significant main effect of sound ((F(1.397,17)=30.175, p=0.000) was found. Restaurant sound was significantly different from traffic (MD=9.4, 95% CI=(6.7-12.0)) and ventilation sounds (MD=7.6, 95% CI=(3.2-12.0)). Traffic and ventilation sounds were not significantly different (Figure 3).

A significant main effect of level ((F(1.190,17)=177.145, p=0.000) was found. All levels were found significantly different from each other (0 dB versus -6 dB: MD=12.6, 95% CI=(9.5-15.7); +6 dB versus -6 dB: MD=26.8, 95% CI=(21.7-32.0); +6 dB versus 0 dB: MD=14.2, 95% CI=(11.5-17.0)) (Figure 3).

A significant two-way interaction was found between technique and sound ((F(2.814,17)=4.398, p=0.009). Restaurant sound was judged louder for BH2, traffic and ventilation sounds were judged louder for ML.

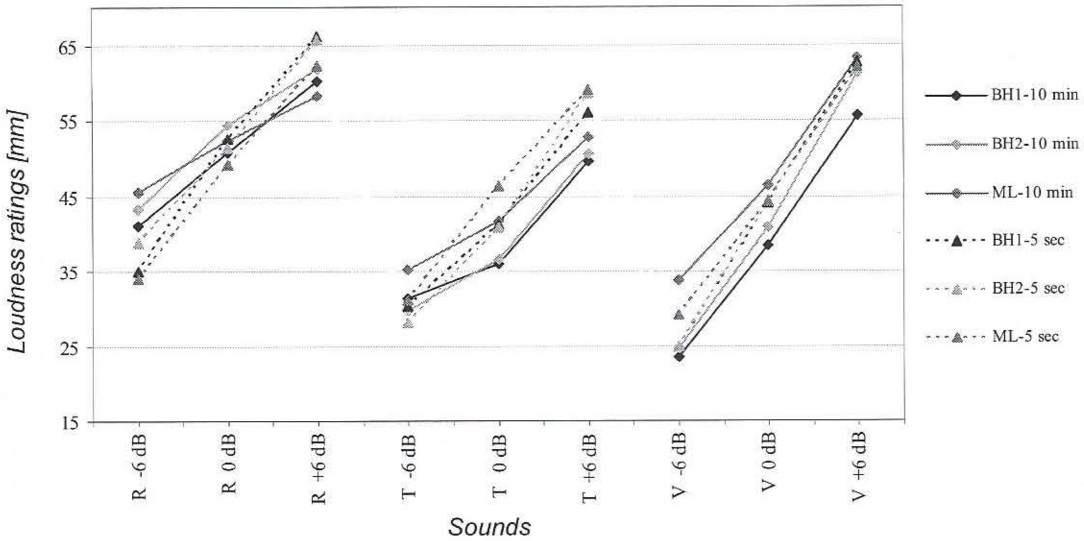

Figure 3 The mean values of loudness ratings for 18 subjects for all combinations of sound, level, technique and duration.

A significant two-way interaction was found between sound and level ((F(4,17)=21.186, p=0.000). Regardless of the sound +6 dB level was always judged louder than 0 dB level which was judged louder than -6 dB level. The difference between levels was higher for ventilation sound.

A significant two-way interaction was also found between level and duration ((F(1.363,17)=6.961, p=0.009).). -6 dB level was judged louder during 10-minute exposure but 0 and +6 dB levels were judged louder during 5-second exposure.

3.3 Unpleasantness

No significant main effects of technique and duration on unpleasantness ratings were found.

A significant main effect of sound ((F(2,17)=6.515, p=0.004) was found. Traffic sound was significantly different from restaurant (MD=-10.2, 95% CI=(-3.3 to -17.1)) and ventilation sounds (MD=-6.4, 95% CI=(-0.2 to -12.6)). Restaurant and ventilation sounds were not significantly different (Figure 4).

A significant main effect of level ((F(1.131,17)=71.757, p=0.000) was found. All levels were found significantly different from each other (0 dB versus -6 dB: MD=7.9, 95% CI=(5.4-10.3); +6 dB versus 0 dB: MD=10.9, 95% CI=(7.1-14.7); +6 dB versus -6 dB: MD=18.8, 95% CI=(13.2-24.5) (Figure 4).

A significant two-way interaction was found between technique and sound ((F(4,17)=4.494, p=0.003). Restaurant sound was judged less unpleasant for ML, traffic sound was judged less unpleasant for BH2 and ventilation sound was judged less unpleasant for BH1 technique.

A significant two-way interaction was found between sound and level ((F(4,17)=3.232, p=0.017). Regardless of the sound +6 dB level was always judged more unpleasant than 0 dB level which was judged more unpleasant than -6 dB level. The difference between levels was higher for ventilation sound.

A significant two-way interaction was found between sound and duration ((F(1.431,17)=26.959, p=0.000). Restaurant sound was judged more unpleasant during 10-minute exposure. Traffic and ventilation sounds were judged more unpleasant during 5 second exposure.

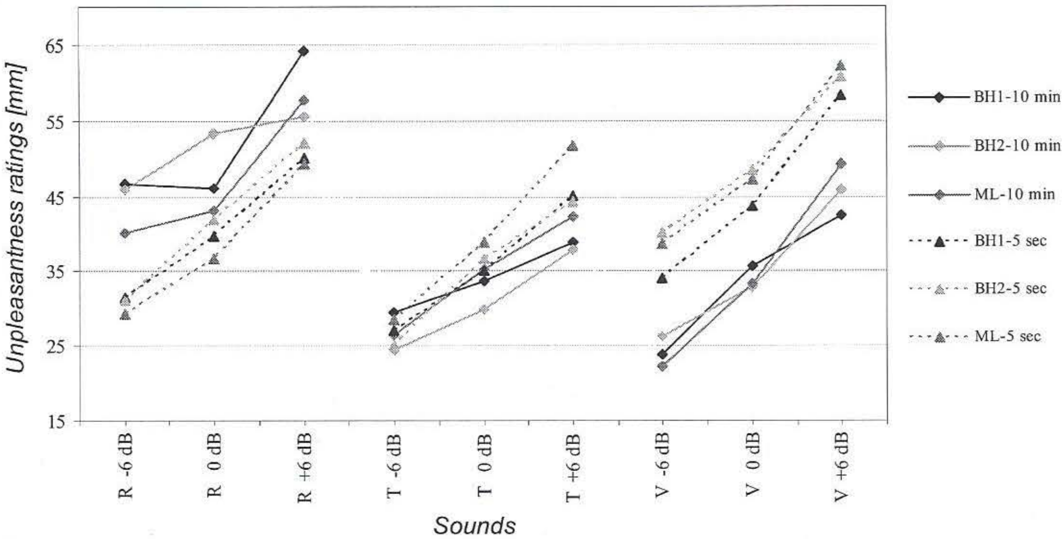

Figure 4 The mean values of unpleasantness ratings for 18 subjects for all combinations of sound, level, technique and duration.

A significant two-way interaction was found between level and duration (($F(2,17)=3.361$, $p=0.047$). -6 dB level was judged equally unpleasant during 10-minute and 5-second exposures but 0 and +6 dB levels were judged louder during 5-second exposure.

3.4 Relations between annoyance, loudness and unpleasantness

Figure 5 show the mean values of 10 minutes annoyance, loudness and unpleasantness judgements for each independent group at 0 dB level. The psychoacoustic attributes are only significantly different for 10-minute ventilation sound when it is played back through open headphone (BH2 technique) (($F(2,53)=4.117$, $p=0.022$). Annoyance judgments of the sounds were found significantly different from loudness judgments (MD=16.3, 95% CI=(4.3-28.2)). No significant difference was found for any of the psychoacoustic attributes during 5-second exposures.

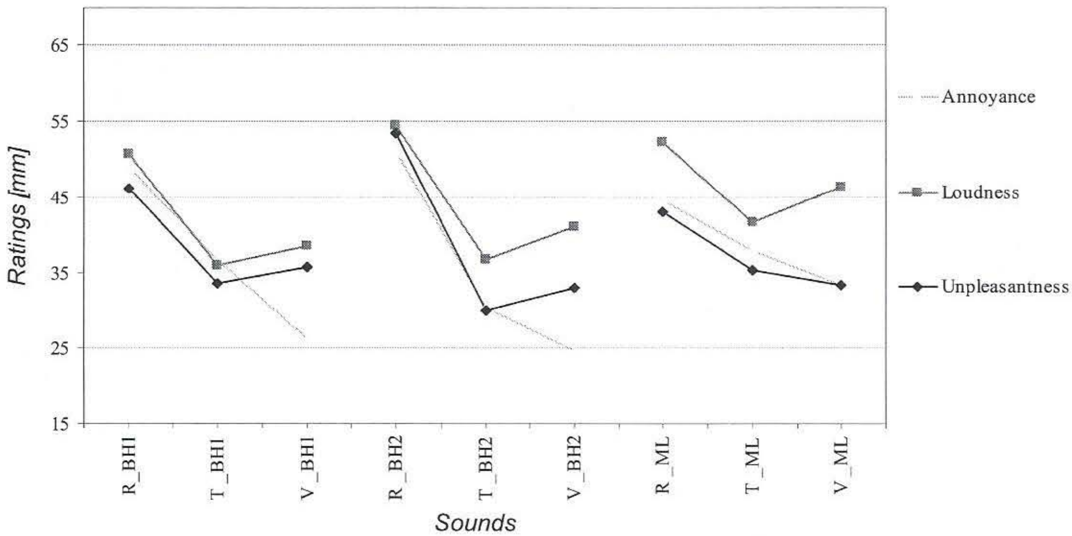

Figure 5 The mean values of 10 minutes for A, L, U ratings at 0 dB level (sounds, techniques).

For each independent group (N=18) the individual paired comparison matrices were pooled across subjects, resulting in the cumulative preference matrix. In this matrix each entry specifies the absolute frequency with which the sound identified by the row of the table was judged as more annoying/loud/unpleasant than the sound identified by the column of the table. Stochastic transitivity checks were performed for each independent group using R 2.0.1 statistical software. The results did not ful-fill the restrictions for the moderate and strong stochastic transitivity which are a prerequisite for an interval and ratio scale. Therefore the data were evaluated with respect to weak stochastic transitivity a prerequisite for an ordinal representation of the data. For each cumulative matrix all the columns of each row was summed. This gives how many times a sound was preferred over the other sounds in the test. The result allows to determine the relative order (ranking) of the sounds (Table 6).

Table 6 The relative order of the sounds both from Method I and II for each group.(the sounds were sorted according to paired comparisons order)

Annoyance				Loudness				Unpleasantness			
Sound	Scale		PC	Sound	Scale		PC	Sound	Scale		PC
	10-m	5-s	5-s		10-m	5-s	5-s		10-m	5-s	5-s
T_m_m	5	4	1	V_m_m	3	3	1	T_m_m	3	9	1
T_h_m	2	3	2	R_m_m	9	5	2	T_h_m	10	3	2
V_m_m	4	2	3	T_m_m	4	4	3	R_m_m	4	2	3
V_h_m	1	1	4	V_h_m	1	1	4	R_h_m	13	4	4
R_m_m	11	7	5	T_h_m	2	2	5	V_m_m	1	7	5
T_m_o	9	9	6	R_h_m	8	6	6	V_h_m	2	1	6
R_h_m	15	10	7	T_m_o	7	10	7	T_m_o	6	13	7
V_m_o	7	5	8	T_h_o	5	7	8	R_m_o	8	8	8
T_h_o	6	8	9	V_m_o	10	9	9	T_h_o	12	6	9
V_h_o	3	6	10	R_m_o	12	11	10	R_h_o	16	10	10
R_m_o	14	11	11	V_h_o	6	8	11	V_m_o	7	12	11
R_h_o	16	16	12	R_h_o	14	12	12	V_h_o	5	5	12
T_m_p	13	12	13	T_m_p	13	14	13	R_m_p	11	15	13
T_h_p	10	13	14	V_m_p	18	16	14	T_m_p	14	17	14
V_m_p	12	14	15	R_m_p	15	17	15	T_h_p	18	14	15
R_m_p	18	17	16	T_h_p	11	13	16	R_h_p	17	16	16
V_h_p	8	15	17	V_h_p	16	15	17	V_m_p	15	18	17
R_h_p	17	18	18	R_h_p	17	18	18	V_h_p	9	11	18

The correlation between rank orders of the three attributes were calculated and showed high correlation (A-L=0.961, A-U=0.930, L-U=0.926). Furthermore, in order to be able to compare the two psychoacoustic methods (scaling and forced choice paired comparisons) the data from Method I were also ranked (Table 6). The correlations between these 9 groups were calculated and Figure 6 represents the multi-dimensional solution of the correlation distance. Scales which are close on the figure have high correlation. Figure 6 indicates that there is a high correlation between 10 minute and 5 second loudness judgments for scaling but the same relation can not be seen for annoyance and unpleasantness.

4. CONCLUSIONS

The results show that for all judgments (annoyance, loudness and unpleasantness), there was no significant main effect of recording and playback techniques; however significant interactions between techniques and sounds were found. For annoyance and unpleasantness, an influence of psychoacoustic method was found.

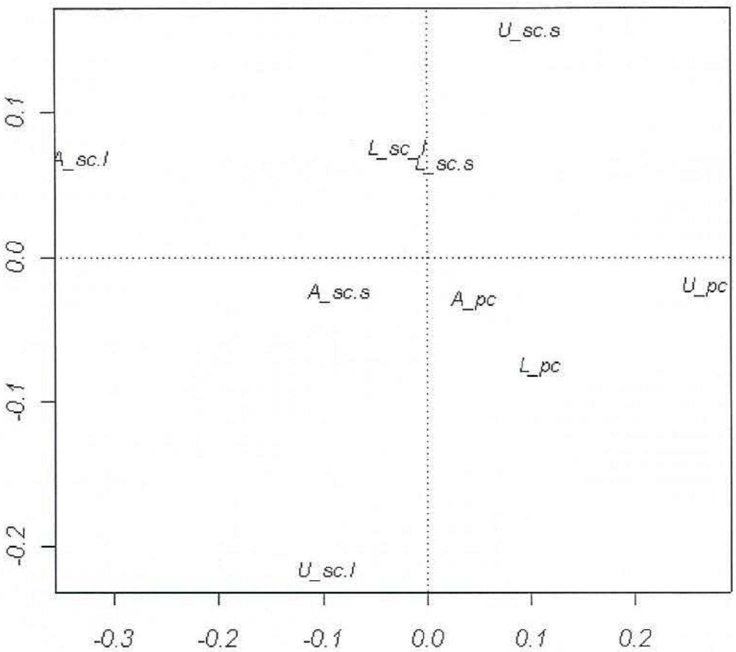

Figure 6 Classical multi dimensional scaling for the rank orders of scaling and paired comparisons data (Group: A=annoyance, L=loudness, U=unpleasantness; Method: sc.l=scale_long, sc.s=scale_short, pc=paired comparisons)

ACKNOWLEDGEMENTS

The project was sponsored by the Danish Technical Research Council. We gratefully acknowledge the assistance of Christian Sejer Pedersen in the preparation of sounds and Steffen Pedersen in the data collection

REFERENCES

[1] J. Blauert, U. Jekosch, Sound quality Evaluation/ A multi layered problem, *Acustica/acta acustica*, **83**, 747-753.

[2] F. Christensen, C. B. Jensen, H. Møller, The design of VALDEMAR - an artificial head for binaural recording purpose, *Proceedings of Audio Engineering Society's 109th Convention*, Los Angeles, California, USA, 2000, pp. 1-8.

[3] U. Landström, S. Arlinger, S. Hygge, Ö. Johansson, A. Kjellberg, K. Persson Waye, Annoying noise. Review for criteria documentation (in Swedish). Arbete och Hälsa. Report 1999:7, National Institute for Working Life, Solna, Sweden

[4] K. Persson Waye, E. Öhrström, Psycho-acoustic characters of relevance for annoyance of wind turbine noise, *Journal of Sound and Vibration* **250**, 65-73, 2002

[5] A. Preis, Measurement of annoyance components, *Internoise 96*, 2289-2294

[6] P. Susini, S. McAdams, S. Winsberg, A multidimensional technique for sound quality assessment, *Acustica/acta acustica*, **85**, 650-656. 1999

[7] N. D. Weinstein, Individual differences in reactions to noise: A longitudinal study in a college dormitory. *J. Appl. Psychol.* **63**, 458-466, 1978.

[8] E. Zwicker, H. Fast, Psychoacoustics facts and models, 1990.